



The Bioinformatics and Biosemiotics of Protein Sorting

Nielsen, Henrik

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Nielsen, H. (2018). *The Bioinformatics and Biosemiotics of Protein Sorting*. Abstract from 18th Annual Biosemiotics Gathering, Berkeley, California, United States.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The Bioinformatics and Biosemiotics of Protein Sorting

Henrik Nielsen

Department of Bio and Health Informatics,

Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Proteins exist and carry out their functions in various compartments of the cell – some even belong *outside* the cell. It is of vital importance to the cell and the organism that the right proteins are transported to their proper compartments after (or during) their synthesis. But what determines this sorting?

In 1999, the Nobel Prize in Physiology or Medicine was awarded to Günter Blobel for the discovery that “proteins have intrinsic signals that govern their transport and localization in the cell.”¹ By this very wording, the Nobel Assembly implicitly acknowledged that a semiotic entity – a signal – can have causal powers in a biological cell. But is this label of “signal” merely a convenient shorthand for complex physico-chemical causal interactions, or is there actual semiosis happening in the recognition of protein sorting signals?

This recognition obviously works in the cell, but it is not easily emulated in the computer. Even the best known sorting signal, the *signal peptide*, which directs its protein to the secretory pathway, is highly variable, making the distinction between signal peptides and other sequences non-trivial. In fact, I have spent most of my career as a bioinformatician trying to recognize signal peptides and other protein sorting signals in amino acid sequences by various machine learning methods. This has resulted in, e.g., the SignalP program for prediction of signal peptides,² which has gathered thousands of citations.

Recently, my group published the DeepLoc program for multi-class prediction of protein subcellular localization based on deep recurrent neural networks.³ In addition to the class prediction, DeepLoc outputs an attention function, showing which parts of the sequence were important for the prediction. I will show how DeepLoc, given only amino acid sequences and class labels during training, is on its own able to locate relevant sorting signals in the sequences.

However, the sorting signals are not like words in a human language, and they are not arbitrary with respect to the physico-chemical properties of the constituent amino acids. I will discuss whether the relationship between a protein sorting signal – viewed as an Aristotelian formal cause – and protein function in the proper compartment – viewed as the corresponding final cause – is indeed semiotic in the context of the living organism.

¹ https://www.nobelprize.org/nobel_prizes/medicine/laureates/1999/press.html

² Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Meth* **8**, 785–786 (2011).

³ Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H. & Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**, 3387–3395 (2017).